

**Ekstraksi Informasi Kesehatan Masyarakat dari Tweet
Berbahasa Indonesia Berbasis Klasifikasi dengan
*Algoritma Naive Bayes***

TUGAS AKHIR

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang**



Oleh :

Khoirir Rosikin

201310370311051

**JURUSAN INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2018**

LEMBAR PENGESAHAN

Ekstraksi Informasi Kesehatan Masyarakat dari Tweet Berbahasa Indonesia Berbasis Klasifikasi dengan *Algoritma Naive Bayes*

TUGAS AKHIR

Sebagai Persyaratan Guna meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Khoirir Rosikin
(201310370311051)

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 2 April 2018

Menyetujui,


Penguji I,



Yuda Munarko, S.Kom., M.Sc.

NIDN: 0706077902

Penguji II,



Vinna Rahmayanti S.N., M.Si.

NIP: 180306071990

Mengetahui,

Ketua Jurusan Teknik Informatika



Gita Indah M., S.T., M.Kom.

NIDN: 0720038101

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

NAMA : KHOIRIR ROSIKIN

NIM : 201310370311051

FAK./JUR. : TEKNIK/INFORMATIKA

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **Ekstraksi Informasi Kesehatan Masyarakat dari Tweet Berbahasa Indonesia Berbasis Klasifikasi dengan Algoritma *Naive Bayes*** beserta seluruh isinya adalah karya saya sendiri bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Setio Basuki, S.T., M.T.
NIDN: 0714028403

Malang, 22 Mei 2018
Yang Membuat Pernyataan



Khoirir Rosikin

LEMBAR PERSETUJUAN

Ekstraksi Informasi Kesehatan Masyarakat dari Tweet Berbahasa Indonesia Berbasis Klasifikasi dengan Algoritma *Naive Bayes*

Khoirir Rosikin

(201310370311051)

Sebagai persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

Malang, 26 Maret 2018

Menyetujui,

Dosen Pembimbing I



Setio Basuki, S.T., M.T.

NIDN: 0714028403

Dosen Pembimbing II



Yufis Azhar, S.Kom., M.Kom.

NIDN: 0728088701

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala limpahan rahmat dan hidayah-Nya, serta salawat dan salam kepada Rasulullah SAW sehingga penulis dapat menyelesaikan tugas akhir yang berjudul:

**“Ekstraksi Informasi Kesehatan Masyarakat dari Tweet
Berbahasa Indonesia Berbasis Klasifikasi dengan
Algoritma *Naive Bayes*”**

Di dalam tulisan ini diuraikan pokok-pokok pembahasan yang meliputi *text mining*, dan juga penerapan algoritma pembelajaran *Naive Bayes* untuk mengklasifikasikan kata dari tweet berbahasa Indonesia.

Penulisan tugas akhir ini dimaksudkan sebagai salah satu syarat kelulusan dari pembelajaran jenjang S1 pada Program Studi Teknik Informatika di Universitas Muhammadiyah Malang.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih banyak kekurangan. Oleh karena itu penulis mengharapkan saran yang membangun agar tulisan ini dapat bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 22 Mei 2018

Penulis

DAFTAR ISI

COVER	
LEMBAR PERSETUJUAN.....	I
LEMBAR PENGESAHAN.....	II
LEMBAR PERNYATAAN.....	III
KATA PENGANTAR.....	IV
ABSTRAK.....	V
ABSTRACT.....	VI
DAFTAR ISI.....	VII
DAFTAR GAMBAR.....	X
DAFTAR TABEL.....	XI
DAFTAR GRAFIK.....	XII
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Batasan Masalah.....	5
1.5 Metodologi.....	5
1.5.1 Studi Literatur.....	5
1.5.2 Desain Eksperimen.....	5
1.5.3 Tahap Implementasi.....	6
1.5.4 Pengujian.....	6
1.5.5 Pembuatan Laporan.....	7
1.6 Sistematika Penulisan.....	7
BAB II.....	9
LANDASAN TEORI.....	9
2.1 Studi Literatur.....	9

2.2 <i>Text Mining</i>	10
2.3 Ekstraksi Informasi.....	11
2.3 Klasifikasi.....	11
2.4 <i>Naive Bayes</i>	13
2.5 Pengujian Confusion Matrix.....	13
BAB III.....	15
ANALISA DAN PERANCANGAN SISTEM.....	15
3.1 Deskripsi Sistem.....	15
3.2 Diagram Alir Eksperimen Sistem.....	15
3.3 Data Penelitian.....	16
3.4 <i>Preprocessing</i> Data.....	17
3.5 Anotasi Data.....	18
3.5 Ekstraksi Fitur.....	20
3.6 Pembentukan Model.....	23
3.7 Klasifikasi.....	23
3.8 Perancangan Pengujian.....	27
BAB IV.....	30
IMPLEMENTASI DAN PENGUJIAN SISTEM.....	30
4.1 Implementasi Kebutuhan Sistem.....	30
4.1.1 Kebutuhan Perangkat Keras.....	30
4.1.2 Kebutuhan Perangkat Lunak.....	30
4.2 Implementasi Perangkat Lunak.....	31
4.2.1 Data Penelitian.....	31
4.2.2 <i>Preprocessing</i> Data.....	31
4.2.3 Anotasi Data.....	33
4.2.4 Ekstraksi Fitur.....	33
4.2.5 Pembentukan Model.....	37
4.2.6 Klasifikasi.....	38
4.3 Pengujian Sistem.....	39

4.3.1 Pengujian Model.....	39
4.3.2 Pengujian Klasifikasi.....	40
BAB V.....	42
PENUTUP.....	42
5.1 Kesimpulan.....	42
5.2 Saran.....	42
DAFTAR PUSTAKA.....	43



DAFTAR GAMBAR

Gambar 2. 1 Proses Klasifikasi.....	11
Gambar 2. 2 Tahapan Klasifikasi.....	12
Gambar 3. 1 Diagram Alir Eksperimen Sistem.....	15
Gambar 4. 1 Data Penelitian.....	31
Gambar 4. 2 Antarmuka Preprocessing Data.....	32
Gambar 4. 3 Hasil Preprocessing Data.....	32
Gambar 4. 4 Hasil Anotasi Data.....	33
Gambar 4. 5 Antarmuka Ekstraksi Fitur.....	34
Gambar 4. 6 Hasil Ekstraksi Fitur.....	34
Gambar 4. 7 Antarmuka Konversi CSV ke ARFF.....	35
Gambar 4. 8 Hasil Ekstraksi Fitur ARFF.....	36
Gambar 4. 9 Antarmuka Pembentukan Model.....	38
Gambar 4. 10 Antarmuka Ekstraksi Informasi Berbasis Klasifikasi.....	39
Gambar 4. 11 Pengujian Model.....	40

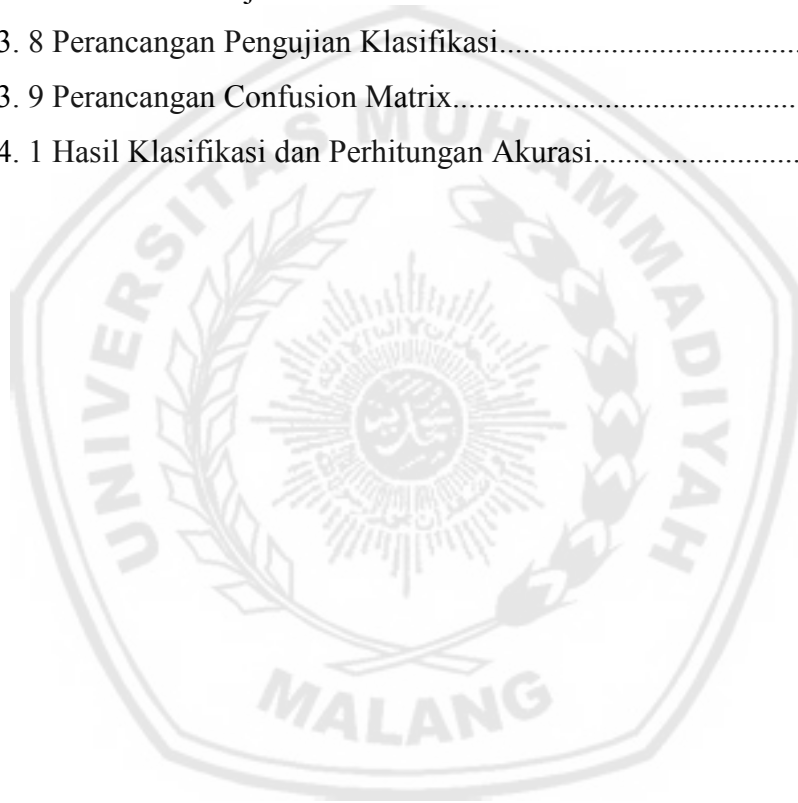
DAFTAR GRAFIK

Grafik 4. 1 Grafik Jumlah Data per Class.....	36
Grafik 4. 2 Grafik Jumlah Data per Class Setelah Resample.....	37



DAFTAR TABEL

Tabel 3. 1 Contoh Data Tweet.....	16
Tabel 3. 2 Label Anotasi Data.....	19
Tabel 3. 3 Contoh Anotasi Data.....	20
Tabel 3. 4 Set Kelas.....	21
Tabel 3. 5 Set Fitur.....	22
Tabel 3. 6 Contoh Ekstraksi Fitur.....	24
Tabel 3. 7 Contoh Data Uji.....	24
Tabel 3. 8 Perancangan Pengujian Klasifikasi.....	27
Tabel 3. 9 Perancangan Confusion Matrix.....	28
Tabel 4. 1 Hasil Klasifikasi dan Perhitungan Akurasi.....	41



DAFTAR PUSTAKA

- [1] “6 Masalah Kesehatan Terbesar di Indonesia.” [Online]. Available: <https://www.guesehat.com/6-masalah-kesehatan-terbesar-di-indonesia>. [Accessed: 25-Mar-2018].
- [2] C.-E. A. Winslow, “The Untilled Fields of Public Health,” *Science* (80-.), vol. 51, no. 1306, pp. 23–33, 1920.
- [3] C. Manning, “Information Extraction and Named Entity Recognition,” p. 73, 2012.
- [4] C. C. Aggarwal and C. X. Zhai, *Mining Text Data*, vol. 9781461432. 2012.
- [5] D. E. Appelt, “Introduction to information extraction,” *AI Commun.*, vol. 12, no. 3, pp. 161–172, 1999.
- [6] A. I. Riaddy, Y. Sibaroni, S. Si, A. Aditsania, S. Si, and M. Si, “Ekstraksi Informasi pada Makalah Ilmiah dengan Pendekatan Supervised Learning,” *e-Proceeding Eng.*, vol. 3, no. 1, pp. 1184–1190, 2016.
- [7] R. Ilyas and M. L. Khodra, “Ekstraksi Informasi 5W1H pada Berita Online Bahasa Indonesia,” vol. 3, no. 1, pp. 35–41, 2015.
- [8] S. Kumar, F. Morstatter, and H. Liu, “Twitter Data Analytics,” *Springer*, p. 89, 2013.
- [9] D. Anggaeska and A. Purwarianti, “Information extraction of public complaints on Twitter text for bandung government,” *Proc. 2014 Int. Conf. Data Softw. Eng. ICODSE 2014*, 2014.
- [10] M. L. Khodra, P. Ayu, A. Insanudin, and M. Megally, “Ekstraksi Informasi Transaksi Online pada Twitter,” *Cybermatika*, vol. 1, no. July, pp. 1–4, 2013.
- [11] L. N. Wulansari, “EKSTRAKSI LOKASI DAN PRODUK DARI DATA TRANSAKSI ONLINE PADA TWITTER,” Universitas Muhammadiyah Malang, 2015.
- [12] A. Hotho, A. Nürnberger, and G. Paass, *A Brief Survey of Text Mining*, vol. 20. 2005.
- [13] A. Z. Arifin and A. N. Novan, “Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass Clustering,” *Pros. Semin. Intell. Technol. its Appl. (SITIA), Tek. Elektro, Inst. Teknol. Sepuluh Nop. Surabaya*, 2002.
- [14] S. Sitanggang, “Pengkategorian Buku untuk Fakultas Teknik menggunakan Metode Naïve Bayes Classifier di Perpustakaan Universitas Widyatama,” p. 2015, 2015.

- [15] F. A. Hermawati, *Data Mining*. 2013.
- [16] P. P. Widodo, *Penerapan Data Mining Dengan MATLAB*. Bandung: Prestasi Pustaka, 2013.
- [17] C. Darwin and T. Origin, *Machine Learning in Action*, vol. 310, no. December. 2005.
- [18] T. M. Mitchell, *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.
- [19] I. H. Witten, E. Frank, and M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. 2011.
- [20] R. T. Pahlevy, "MENENTUKAN PENERIMA BEASISWA DENGAN MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING (SAW)," UNIVER SITAS PEMBANGUNAN NASIONAL, 2011.
- [21] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif, and F. Ferdian, "InaNLP: Indonesia natural language processing toolkit, case study: Complaint tweet classification," *4th IGNITE Conf. 2016 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2016*, pp. 5–9, 2016.
- [22] T. R. Hoens and N. V Chawla, "Imbalanced Datasets: From Sampling to Classifiers," *Imbalanced Learn.*, pp. 43–59, 2013.